A Framework for Automatic Acquisition of Croatian and Serbian Verb Aspect from Corpora

Samardžić, Tanja ; Miličević, Maja

A Framework for Automatic Acquisition of Croatian and Serbian Verb Aspect from Corpora

Tanja Samardžić[1], **Maja Miličević**[2]
[1]Corpus Lab, University of Zurich, URPP Language and Space, Freiestrasse 16, CH-8032 Zurich, Switzerland
[2]Faculty of Philology, University of Belgrad, Studentski trg 3, 11000 Belgrade, Serbia
tanja.samardzic@uzh.ch, m.milicevic@fil.bg.ac.rs

Abstract

Verb aspect is a grammatical and lexical category that encodes temporal unfolding and duration of events described by verbs. It is a potentially interesting source of information for various computational tasks, but has so far not been studied in much depth from the perspective of automatic processing. Slavic languages are particularly interesting in this respect, as they encode aspect through complex and not entirely consistent lexical derivations involving prefixation and suffixation. Focusing on Croatian and Serbian, in this paper we propose a novel framework for automatic classification of their verb types into a number of fine-grained aspectual classes based on the observable morphology of verb forms. In addition, we provide a set of around 2000 verbs classified based on our framework. This set can be used for linguistic research as well as for testing automatic classification on a larger scale. With minor adjustments the approach is also applicable to other Slavic languages.

Keywords: verb aspect, event, automatic acquisition, derivational morphology, morphological segmentation, Croatian, Serbian

1. Introduction

Verb aspect is a grammatical and lexical category that encodes temporal unfolding and duration of events described by verbs. It represents a potentially interesting source of information for applications that involve event analysis, such as ordering of events in text summarisation, or identifying whether a relation holds in a particular moment in relation extraction. Despite extensive and long lasting discussion in theoretical linguistics, verb aspect is considerably underexplored as a means of event analysis in computational approaches, partly due to the fact that theoretical accounts are difficult to implement in automatic analysis, and partly because aspectual distinctions are not easily observable in the most widely studied languages (such as English).

Unlike in most other European languages, in the Slavic language family verb aspect is observable through lexical derivations. Consider, for example, the Croatian/Serbian sentences and their translations in (1)-(2).[1] The English past tense form *cooked* is ambiguous, potentially describing a habitual activity, as in (1), or a single completed event, as in (2). The context does sometimes allow us to disambiguate between the two interpretations (see e.g. the modifier *often* in (1)), but such clues are frequently absent (as in (2)). In Croatian and Serbian, the habitual meaning in (1) is expressed with one verb (*kuhao*), and the single completed event in (2) with another (*prokuhao*).

	Vinston	je	često	kuha-o.
(1)	Winston	AUX	often	cooked.
	Winston often cooked$_I$.			

	Vinston	je	pro-kuha-o	čašu	vode.
(2)	Winston	AUX	coooked	glass	water.
	Winston boiled$_P$ *a glass of water.*				

The two verbs are morphologically related. The presence of a prefix (*pro-*) in *prokuhao* indicates that the described event is completed; i.e., verb aspect is encoded via a lexical derivation. However, this encoding is not direct, as similar verb affixes are not formal aspect markers. The prefix *pro-* in (2), for instance, is similar in meaning to the English preposition 'through', and it contributes this lexical content to the verb, while at the same time changing the verb's aspect. Other prefixes (e.g. *u-* 'in', *od-* 'from') can have a similar effect on verb aspect. However, aspect can also be assigned independently of the presence of a prefix (see Section 3.1.). Aspect encoding through lexical derivation is thus highly ambiguous and inconsistent, presenting numerous challenges for generalisation. As a consequence, Slavic verb aspect is usually considered to be highly complex, involving an interaction of semantic and formal categories that has not yet been fully understood.

In this paper, we propose a novel framework for classifying Croatian and Serbian verb types into a number of relatively fine-grained aspectual categories based on the observable verbal morphology. Unlike most theoretical accounts, we offer a data-driven approach: each aspectual category is associated with a particular combination of verb affixes. Our framework is based on theoretical insights, but its main purpose is to facilitate computational processing. It is intended to serve in automatic recognition of event properties such as duration (short or long), completeness (having an end point or not), and dynamics (involving a change of state or not), which are indirectly expressed by verb aspect. In addition to the proposed data-driven categories, we provide a set of around 2000 verb types with manually assigned aspectual categories based on our framework. This set can be used for quantitative linguistic research and for evaluation of automatic classification on a larger scale. We focus on data

[1]Despite being two separate standard languages, both Croatian and Serbian are discussed in the paper, as their close relatedness would make separate treatment highly inefficient. All examples in the paper are well-formed sentences in both languages (with minor adjustments such as *kuhati* [Croatian] vs. *kuvati* [Serbian]).

from Croatian and Serbian, but with minor adjustments the approach is also applicable to other Slavic languages.

2. Related Work

A number of aspectual categories have been proposed in the literature to describe the full range of interpretations that can be assigned to the events described by verbs. Two main categories are universally identified as grammatical aspect in Slavic languages: *imperfective* (illustrated by the sentence in (1)), and *perfective* (illustrated in (2)). Imperfective verbs approximately correspond to temporally unbounded, and perfective verbs to temporally bounded events in terms of Dowty's definition (Dowty, 1979).[2] This mapping does not apply to all cases, but we refer to it as the closest general description of the two categories. In addition to these two main categories, grammars occasionally refer to notions such as *iterative* (repeated event or activity, e.g. turning pages), *semelfactive* (instantaneous event, e.g. clicking on a web page), or *biaspectual* verbs (ambiguous between perfective and imperfective, e.g. calling somebody) (Stanojčić and Popović, 1995). Grammars typically do not address the relations between the categories.

Theoretical accounts offer different frameworks for organising aspectual categories based on more general underlying principles. Our work is most closely related to those approaches that address empirical issues such as observable morphology and verb classification. For instance, Janda (2007) proposes a semantic space consisting of three dimensions: *open vs. close*, *completable vs. non-completable*, *durative vs. instantaneous*. This space is then used to define four types of perfective verbs in Russian: natural perfective, specialised perfective, single act, complex acts; verbs are grouped into clusters based on the perfective types that they exhibit. Sonnenhauser (2006) and Zangenfeind and Sonnenhauser (2014) propose six aspect types, three perfective and three imperfective, based on the temporal scope of the event described by the verb with respect to a broader context (termed *topic*).

Certain relations between morphology and aspect interpretation are invoked in both approaches. Janda (2007) notes the relationship between prefixes and perfectivity and suffixes and some other aspect categories. Sonnenhauser (2006) and Zangenfeind and Sonnenhauser (2014) argue for a one-to-one mapping between aspect types in Russian verbs and English past tenses. Our study pursues the question of morphological marking and aspect types further. Our framework does not follow any proposed typology of verb aspect entirely. We make use of the most general notions (perfective and imperfective), but we redefine them in terms of observable morphological features.

Focusing on the structural side of verb aspect, our approach assumes the general structure of Slavic verb derivations developed in the framework of Generative Grammar (Svenonius, 2004a; Di Sciullo and Slabakova, 2005; Arsenijević, 2006), illustrated by the embeddings in (3).

(3) [[[prefix [stem]] suffix]

[2] Although imperfective and perfective aspect are grammatical categories in Slavic languages, they are not easily distinguishable from inner (or lexical) aspect discussed by Dowty (1979).

We follow Arsenijević (2007) and Žaucer (2010) in that we do not distinguish between lexical and superlexical prefixes argued for by Svenonius (2004b) and Milićević (2004).

On the side of natural language processing, our study is set in the context of work on derivational morphology (Zeller et al., 2013; Šnajder, 2014). Unlike previous approaches that mostly address changes in category, our work is focused on derivations within the category of verbs. In addition to establishing a derivational relation between verbs, we aim at defining the kind of relation, i.e. at checking how the derivation affects verb aspect. The framework proposed in our study is suitable for automatic extraction based on morphological segmentation (Ruokolainen et al., 2013; Cotterell et al., 2015).

3. Verb Aspect and Morphology

With some exceptions (Samardžić and Miličević, 2013), most accounts of verb aspect in Croatian and Serbian rely on the traditional notion of aspectual verb pairs (Mrazović and Vukadinović, 2009; Jelaska and Opačić, 2005), which is largely abandoned in more recent theory. Regarding verbs as aspectual pairs accounts for contrasts such as the ones illustrated in (1)-(2). This view, however, fails to take into account the fact that multiple pairs share the same base verb, whose meaning and aspect are modified in different ways by different affixes. Consider the examples in (4)-(5).

(4)
Vinston je kuh-nu-o malo vode.
Winston AUX cooked little water.
Winston boiled$_P$ a bit of water.

(5)
Vinston je pro-kuha-va-o čašu vode
Winston AUX cooked glass water
(kada je čuo glas).
(when AUX heard sound).
Winston was boiling$_I$ a glass of water when he heard the sound.

The verbs in (4)-(5) do not constitute a typical aspectual pair, although one is perfective and the other imperfective. Moreover, these verbs are morphologically related to those in (1)-(2).

To account for these facts, we argue that Croatian and Serbian aspectual verb derivations can be better described as a more complex structure, where a number of verbs are nested around a single base verb. This nesting takes the form of a matrix, where groupings in one dimension are based on prefixation, and those in the other on suffixation. The groupings based on prefixation, represented as the vertical dimension in our matrices, correspond to clustering proposed by Janda (2007). The horizontal, suffixation dimension corresponds to the derivations discussed in the framework of Generative Grammar. Combining the two dimensions results in relatively sparse matrices which can be populated in a more or less regular fashion.

3.1. Verb Aspect Matrices

Regular aspectual sequences revolve around a base verb, which is typically imperfective, but can also be perfective. A similar basic set of derivations is allowed in both cases, as shown in Tables 1 and 2.

x=kuhati 'cook$_I$'		
With prefix		No prefix
pref-(x)$_P$	**pref-(x)-suff**$_I$	**(x)-suff**$_P$
s-kuhati 'cook'	—	kuh-**nu**-ti 'cook briefly'
pro-kuhati 'boil briefly'	**pro**-kuha-**va**-ti	
is-kuhati 'cook well'	**is**-kuha-**va**-ti	
ot-kuhati 'clean by boiling'	**ot**-kuha-**va**-ti	
za-kuhati 'add something into boiling liquid'	**za**-kuha-**va**-ti	

Table 1: Aspectual derivations of an imperfective verb

x=baciti 'throw$_P$'		
With prefix		No prefix
pref-(x)$_P$	**pref-(x)-suff**$_I$	**change(x)**$_I$
pre-baciti 'throw over'	**pre**-baci-**va**-ti	bac**a**ti 'throw$_I$'
iz-baciti 'throw out'	**iz**-baci-**va**-ti	
u-baciti 'throw in'	**u**-baci-**va**-ti	
od-baciti 'throw away'	**od**-baci-**va**-ti	
za-baciti 'throw back'	**za**-baci-**va**-ti	

Table 2: Aspectual derivations of a perfective verb

The perfective verbs in the first column result from prefixation. Most of them can be further modified by a suffix, forming derived imperfectives shown in the second column in both tables. Note that adding a prefix to a verb that is already perfective, as in Table 2, does not result in a change of aspect.

An additional derivational step, not shown in the tables, is possible in both cases: a second (or even a third) prefix can be attached to the derived imperfective changing its aspect back to perfective (e.g. *poizbacivati* 'throw out multiple persons/things'). We omit this case from the schema because such forms are very rare, and thus not likely to be found in a corpus.

Most prefixes in the first column are associated with lexical content, which modifies the meaning of the base verb. Possible combinations of verbs and prefixes are determined by their lexical compatibility, or lexical preferences. Some base verbs can be combined with many different prefixes, while others with only few (see Section 4.2.).

The suffixes, in contrast to the prefixes, are functional morphemes with no lexical content to contribute to the derived verb. Their form might vary, but this variation is not influenced by the verbs' or the suffixes' lexical content. For example, the imperfective suffix can take the form *-ja* instead of *-va* in some verbs (e.g. *napi-**ja**-ti* < *na-piti* < *piti* 'drink'), but this does not result in a different interpretation. Also, the variation is not possible with the same base (if *napi-**ja**-ti* is possible then **napi-**va**-ti* is not possible).

The distribution of suffixes depends on the presence of the prefix (not its identity): imperfective suffixes such as *-va* and *-ja* can only be attached to the verbs that already have a prefix.

Both imperfective and perfective base verbs can be modified without prefixation, which results in eliminating the vertical dimension in the matrices (see the right-hand side of Tables 1 and 2). Changing grammatical aspect from imperfective to perfective is realised through the suffix *-nu*, as shown in Table 1. The suffix *-va* cannot be attached to the perfectives formed using *-nu* to form derived imperfectives. Combining such a suffix with a prefix (an option not shown in the table) is possible, meaning that the vertical dimension of lexical derivation is kept, but such combinations are very marginal (e.g. *prokuhnuti* 'boil very briefly'). The same applies to adding this suffix to the perfective base verb (also not represented in the tables). We come back to the the cases where such forms are not marginal later. On the other hand, changing grammatical aspect from perfective to imperfective without a prefix is realised through various changes in the verb stem. The $i > a$ change shown on the right-hand side of Table 2 is just one of many available options, which are highly irregular.

In summary, it is evident from Tables 1 and 2 that there are two kinds of derivations: with and without prefixes. Prefixation results in a matrix of regular derivations, while the derivations that involve the other side of the stem tend to be one-dimensional and irregular. Further, based on the position of the verb forms in the structure, we can distinguish three kinds of imperfectives: base imperfectives with no affixes, regular derived imperfectives with a prefix and a suffix, and irregular derived imperfectives without a suffix. There are also three kinds of perfective verbs: base perfectives with no affixes, those formed with a prefix, and those formed with a suffix.

3.2. Morphology and Meaning

Each of the morphological verb types can be assigned an aspect-related interpretation, as is evident from the glosses in the examples above. We summarise the kinds of meaning that are typically expressed by the different morphological types in Table 3.

Base imperfective	Habitual, descriptive activities
Regular derived imperfective with *-va*	Particular activities, anchored in time and space
Irregular derived imperfective	Particular activities, less specified than regular imperfectives
Base perfective	A single completed event, underspecified
Prefixed perfective	Completion, start, or some other component of an event
Suffixed perfective	A single instantaneous event

Table 3: Relationship between morphological verb types and aspectual meanings.

The mapping between morphology and meaning is based

on the idea that the presence of a prefix restricts the meaning of a verb. For example, *izbacivati* in Table 2 is more specified than *bacati*, with more restricted selection preferences and complementation patterns, because it contains a prefix.
The mappings presented in Table 3 should not be seen as strict rules. However, locating a particular verb type in a verb derivation matrix is a useful indicator of the properties of the event it describes.

3.3. Sparseness and Irregularity

As already mentioned above, the derivation matrices are not always fully populated. One case of an unfilled position is illustrated in Table 1, where one of the prefixed verbs (*skuhati* 'cook') does not allow the formation of a derived imperfective (**skuhavati*). Similar cases are rather frequent across the verb system (e.g. *pisati* - *napisati* - **napisivati* 'write', *čistiti* - *očistiti* - **očišćavati* 'clean', *čitati* - *pročitati* - **pročitavati* 'read'). However, there is no obvious regularity as to which prefixes behave in this way.[3] Another position that is often not filled is that of the base verb itself, as in the case of frequently used verbs *premestiti*, *izmestiti* 'move', which should both be derived from the non-existent base verb **mestiti*.
The main source of irregularity are the perfective base forms, which tend to have imperfective counterparts derived in not fully transparent ways. Table 2 illustrates imperfective formation through a change in stem vowel (*baciti* > *bac**a**ti*). Some verbs add the suffix *-va* despite not having a prefix, as in *dati* > *da-**va**-ti* 'give', or they use two imperfective suffixes, as in *desiti (se)* > *deš-**a-va**-ti (se)* 'happen' (the first suffix being an allomorph of *-ja*). A group of verbs that contain the suffix *-nu*, but that do not seem to be derived, behaves in a similar way as base perfectives (e.g. *dah-**nu**-ti* 'breathe', *klik-**nu**-ti* 'click', *trep-**nu**-ti* 'blink'). It is in this group that we find most of the non-marginal forms containing both a prefix and the suffix *-nu* (e.g *do-dir-**nu**-ti* 'touch', ***o**-kre-**nu**-ti*, 'turn', ***uz**-dah-**nu**-ti* 'take a breath'), contrary to the general pattern shown in Table 1.
In the following cases, ambiguity can pose problems for the automatic recognition of morphological verb types:

- Forms with no affixes are ambiguous between base imperfectives and base perfectives. The disambiguation would have to rely on the information from the context or morpho-syntactic annotation.
- In some cases it is difficult to distinguish between two prefixes (e.g. *odrati* 'skin' < *o* + *derati*, *odraditi* 'finish doing' < *od* + *raditi*), or between a prefix and a stem part (e.g. *s* in ***s**kuhati* 'cook' vs. *skočiti* 'jump')
- Cases of base homonymy need to be distinguished (e.g. *biti* 'be' vs. 'beat'), and unrelated forms need to be recognised as such (e.g. *pisnuti* 'peep' is not derived from *pisati* 'write')

In these cases, relatedness between derivations cannot be established based on a simple set of rules. However, the fact that related forms tend to be morphologically similar and carry similar meanings constitutes a sound basis for various stochastic approaches involving machine learning.

[3]This issue was addressed in the literature leading to a hypothesised division between lexical and supralexical prefixes (Svenonius, 2004b), which was later shown to be inadequate.

4. The Evaluation Set

In addition to the described framework, we construct an evaluation set that consists of all verbs occurring in the Serbian translation of the novel "1984" by George Orwell, taken from the MULTEXT-East project (Krstev et al., 2004; Erjavec, 2010).[4]
Constructing the test set allowed us to assess the coverage and adequacy of our framework. While we were able to place all verbs in one of the positions in the proposed matrices, we could not always do this based solely on the observable morphology. The cases where the morphological patterns were not sufficient for identifying aspectual categories are described in Section 3.3. above.
The constructed data set is freely available for research purposes. It is intended to be used as a gold standard for evaluating automatic approaches to the extraction of aspectual features from corpora, but also for quantitative linguistic studies of verb aspect and morphological derivations. The data will be distributed through the ReLDI infrastructure (Samardžić et al., 2015) (`https://reldi.spur.uzh.ch`). In this section, we describe the construction of the data set and provide its statistical summary.

4.1. Method

In the lemmatised verb list extracted from the corpus we manually identify the base aspectual forms and assign to each base form all related aspectual derivations occurring in the sample (e.g. *ploviti* 'sail': *doploviti* 'sail to', *proploviti* 'start sailing', *uploviti* 'sail in'). We also encode the relation type (perfective prefixation, perfective suffixation, regular imperfective formation, irregular imperfective formation) and the information on each form's frequency. Base forms that do not occur in the sample are listed with a frequency of 0.
The decisions made with regard to problematic cases are the following:

- We rely on surface morphology and do not attempt to separate cases of polysemy and homonymy (e.g. *kupiti* 'buy' / 'gather' is assigned as the base verb for derivatives semantically related to either meaning); forms derived from non-base imperfectives are assigned to the base perfective (e.g. *zalupati* 'start banging' to *lupiti* 'hit')
- We separate suppletive forms and do not explicitly relate them to each other (e.g. *reći* and *govoriti* 'say' are treated as two different base verbs)
- We treat the base forms absent from the contemporary language (e.g. **mestiti* as a base for *namestiti* 'adjust')

[4]Our original intention was to create the test sets for both Croatian and Serbian, given that the Croatian version of the corpus is listed as available through the Metashare platform (`http://meta-share.ffzg.hr/repository/search/?q=corpus`). Unfortunately, we did not get access to Croatian data.

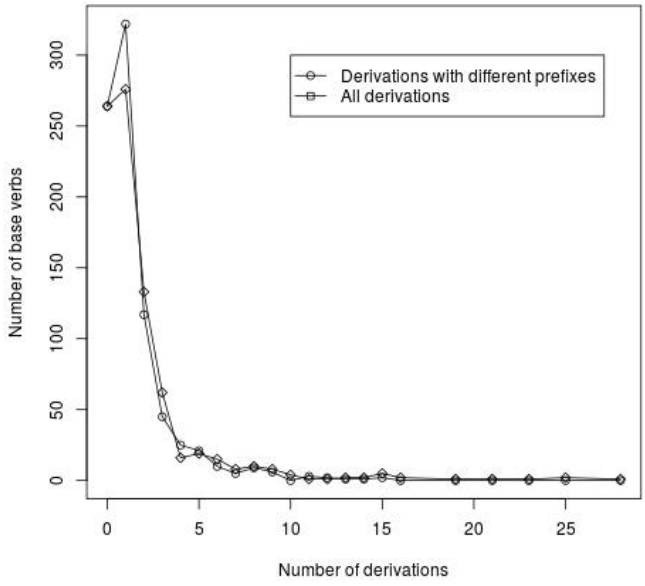

Figure 1: The distribution of associated derivations over base forms in the MULTEXT-East sample.

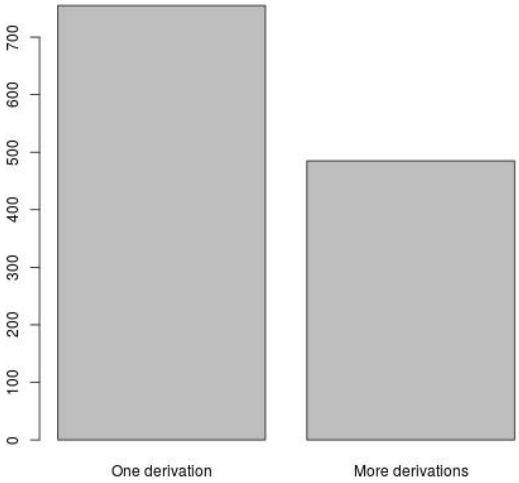

Figure 2: The distribution of derivations over form complexity.

equally as those that do exist, but do not occur in the sample (e.g. *tovariti* 'load', from which *istovariti* 'unload' is derived)

4.2. Data Summary

The total of 2101 verb types are divided into 834 groups, where each group is associated to (and identified with) a unique base form. Figure 1 shows two distributions of aspectual derivations over the base forms: 1) the number of different prefixes associated with the base forms and 2) the total number of derivations. We see that both distributions are skewed to the left, but do not entirely follow the Zipfian line characteristic of much of linguistic data.

Looking at the number of different prefixes associated with base forms (corresponding to the number of rows in Tables 1 and 2), the most frequent case is an association of only one prefix with one verb (322/834). As the number of prefixes grows, the number of base verbs appearing with that number of prefixes drops rapidly. The maximum number of different prefixes associated with a base verb in our sample is 18; two verbs with this number of prefixes are found: *ići* 'go' and *stati* 'stop, stand, fit'.

Focusing on the total number of derivations, the most frequent case is an association of a single derivation with a single base form; in other words, most verb groups in our sample are constituted by verb pairs. Such grouping is observed in around one third of all cases (276/834). Base forms with no derivations (single-verb groups) account for almost as many cases (264/834); particularly numerous among them are verbs of foreign origin (*analizirati* 'analyse', *inficirati* 'infect', etc.). As for the remaining groups (313/834), around 140 verbs have two, around 60 verbs three, and around 100 verbs four or more derivatives. The maximal number of derivations associated with a single base in our sample is 28, for the verb *ići* 'go'.

Comparing the two distributions in Figure 1, we see that the biggest difference between them is in their peaks: more verbs are found that are associated with a single prefix, than with a single derivation. This means that many of the verbs associated with a single prefix allow for further derivations based on the one prefixed form.

Adding to the picture the suffixation dimension of the matrices, Figure 2 shows the number of simple vs. complex derivations. This number refers to how populated the rows in the Tables 1 and 2 are, i.e. how many columns are present for each row. It can be seen that around 60% of the rows (755/1240) contain only one form, typically resulting from prefixation rather than suffixation.

Our data summary shows only the most general tendencies in the data set. Many other relations and other phenomena remain open for future investigations, such as productivity of derivations, dependency relations between different categories (e.g. prefixes and regularity), etc.

5. Conclusion

In this paper, we argue for a data-driven analysis of aspectual verb derivations in Croatian and Serbian, extendible to other Slavic languages. We provide a data set in which we implement the proposed analysis on a sample of around 2000 verbs. While the data show that pair-wise aspectual groupings are frequent, there are also indications of a substantial number of base verbs being better captured through more complex aspectual matrices. The proposed data set constitutes a basis for scaling up the data-driven analysis of verb aspect, which is needed for a better understanding of linguistic encoding of time.

6. Acknowledgement

The research leading to these results was conducted as part of Regional Linguistic Data Initiative, an institutional partnership between the Universities of Zurich, Belgrade and Zagreb, within the Scientific Co-operation between Eastern Europe and Switzerland programme (project No. 160501).

7. Bibliographical References

Arsenijević, B. (2006). *Inner aspect and telicity: The decompositional and the quantificational nature of eventualities at the syntax-semantics interface*. LOT, Utrecht.

Arsenijević, B. (2007). Slavic verb prefixes are resultative. *Cahiers Chronos*, 17:197–213.

Cotterell, R., Müller, T., Fraser, A., and Schütze, H. (2015). Labeled morphological segmentation with semi-markov models. In *Proceedings of the Nineteenth Conference on Computational Natural Language Learning*, pages 164–174, Beijing, China, July. Association for Computational Linguistics.

Di Sciullo, A. M. and Slabakova, R., (2005). *Perspectives on Aspect*, chapter Quantification and Aspect, pages 61–80. Springer Netherlands, Dordrecht.

Dowty, D. R. (1979). *Word meaning and Montague grammar: the semantics of verbs and times in generative semantics and in Montague's PTQ*. D. Reidel, cop., Dordrecht, Boston.

Erjavec, T. (2010). MULTEXT-East version 4: Multilingual morphosyntactic specifications, lexicons and corpora. In Nicoletta Calzolari, et al., editors, *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*, pages 2544–2547, Valletta, Malta. European Language Resources Association (ELRA).

Janda, L. A. (2007). Aspectual clusters of russian verbs. *Studies in Language. International Journal sponsored by the Foundation "Foundations of Language"*, 31(3):607–648.

Jelaska, Z. and Opačić, N. (2005). Glagolski vid i vidski parovi. In Zrinka Jelaska et al., editor, *Hrvatski kao drugi i strani jezik*, pages 152–170. Hrvatska sveučilišna naklada, Zagreb.

Krstev, C., Vitas, D., and Erjavec, T. (2004). MULTEXT-East resources for Serbian. In *Proceedings of 8th Informational Society - Language Technologies Conference, IS-LTC*, pages 108–114, Ljubljana, Slovenia.

Milićević, N. (2004). The lexical and superlexical verbal prefix iz- and its role in the stacking of prefixes. *Nordlyd*, 32(2):279–300.

Mrazović, P. and Vukadinović, Z. (2009). *Gramatika srpskog jezika za strance*. Izdavačka knjižarnica Zorana Stojanovića, Novi Sad.

Ruokolainen, T., Kohonen, O., Virpioja, S., and Kurimo, M. (2013). Supervised morphological segmentation in a low-resource learning setting using conditional random fields. In *Proceedings of the Seventeenth Conference on Computational Natural Language Learning*, pages 29–37, Sofia, Bulgaria, August. Association for Computational Linguistics.

Samardžić, T. and Miličević, M. (2013). Constructing a learner-friendly corpus-based dictionary of serbian verbal aspect. *Primenjena lingvistika*, 14:77–89.

Samardžić, T., Ljubešić, N., and Miličević, M. (2015). Regional Linguistic Data Initiative (ReLDI). In *The 5th Workshop on Balto-Slavic Natural Language Processing*, pages 40–42, Hissar, Bulgaria, September. INCOMA Ltd. Shoumen, Bulgaria.

Sonnenhauser, B. (2006). *Yet There's Method In It Semantics. Pragmatics and the Interpretation of the Russian Imperfective Aspect*. Verlag Otto Sagner, Munich.

Stanojčić, Z. and Popović, L. (1995). *Gramatika srpskoga jezika*. Zavod za udžbenike i nastavna sredstva, Beograd.

Svenonius, P. (2004a). Slavic prefixes and morphology. *Nordlyd*, 32(2):177–204.

Svenonius, P. (2004b). Slavic prefixes inside and outside VP. *Nordlyd*, 32(2):205–253.

Žaucer, R. (2010). The reflexive-introducing na- and the distinction between internal and external Slavic prefixes. In Anastasia Smirnova, et al., editors, *Formal Studies in Slavic Linguistics*, pages 54–102, Newcastle upon Tyne. Cambridge Scholars Publishing.

Zangenfeind, R. and Sonnenhauser, B. (2014). Russian verbal aspect and machine translation. In V. P. et al. Selegej, editor, *Conference Proceedings of Computational Linguistics and Intellectual Technologies / Komp'juternaja lingvistika i intellektual'nye tehnologii*, volume 13 (20), pages 700–709, Moscow.

Zeller, B., Šnajder, J., and Padó, S. (2013). Derivbase: Inducing and evaluating a derivational morphology resource for german. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1201–1211, Sofia, Bulgaria, August. Association for Computational Linguistics.

Šnajder, J. (2014). DerivBase.hr: A high-coverage derivational morphology resource for Croatian. In Nicoletta Calzolari (Conference Chair), et al., editors, *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, Reykjavik, Iceland, may. European Language Resources Association (ELRA).